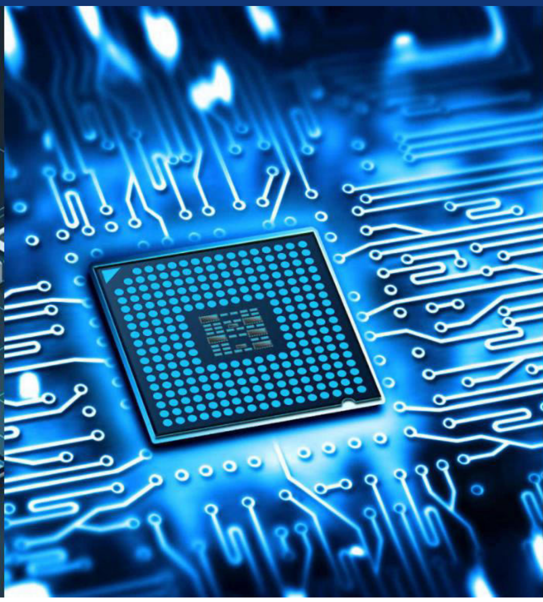
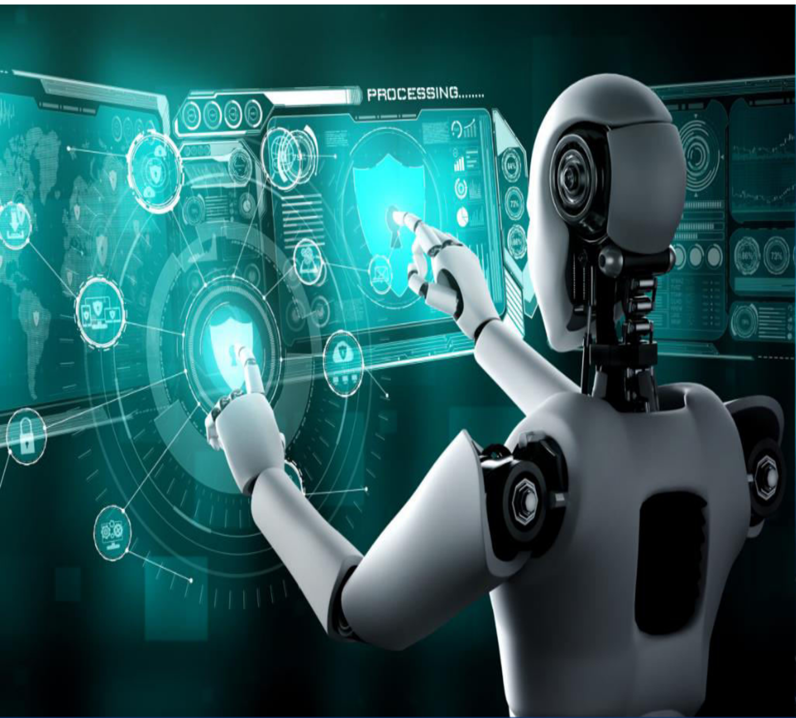
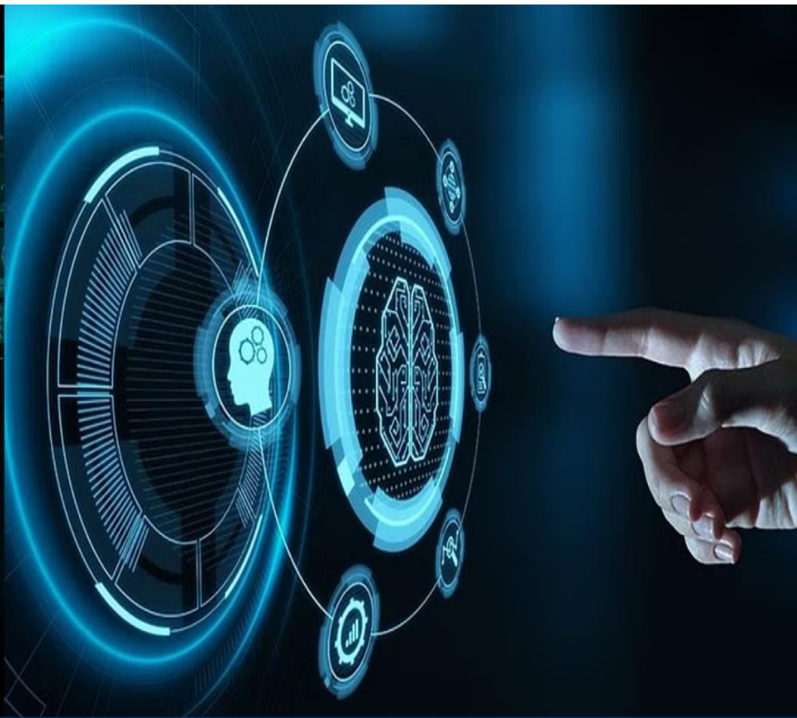
ISSN(O): 2320-9801
ISSN(P): 2320-9798


International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)





Impact Factor: 8.771

Volume 13, Issue 3, March 2025

www.ijircce.com ijircce@gmail.com +91-9940572462 +91 63819 07438

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Enhancing Customer Segmentation in CRM: Integrating Geographic Population Insights with RFM Models

Om .B. Mistry, Mr. Utpal .B. Patel

Department of Computer Science and Engineering, Parul Institute of Technology, Parul University, Gujarat, India

Assistant Professor, Department of Computer Science and Engineering, Parul Institute of Technology, Parul University, Gujarat, India

ABSTRACT: Customer segmentation is an important part of customer relationship management (CRM), allowing companies to better track customer behavior over time. Quality work is essential for businesses with large customer bases. Traditionally, CRM systems have used the Recency, Frequency, and Monetary (RFM) model for segmentation. However, traditional RFM models ignore the impact of location on customer behavior. Today's applications focus primarily on the Recency, frequency, and financial impact of customer interactions. Many studies have investigated the use of unsupervised machine learning algorithms (e.g. K-Means) for customer segmentation using RFM models. However, this model is still insufficient as it ignores other important factors related to the field of application. In this study, we present an improved model by introducing a new variable, P-Score, which integrates insights into the purchasing behavior of a geographic population and population purchase power. These improved models increase the accuracy of predicting customer behavior, allowing the company to identify customers who are likely to respond positively. Our findings based on real e-commerce data highlight the importance of population segmentation in CRM strategies, helping to improve customer retention and accurately predict revenue.

KEYWORDS: Customer Classification, Customer Relationship Management (CRM), Customer Response Prediction, Customer Segmentation, Deep Neural Networks (DNN), Diversity, K-Means Algorithm, Machine Learning, RFM and RFM-D Models, Targeting Optimization.

I. INTRODUCTION

In the world of consumer analytics, getting the gist of purchasing behavior is essential for businesses to make well informed decisions about marketing, sales, and customer engagement. One of the best and popular ways to analyze customer behavior is through RFM analysis, a method that identifies customers based on their Recency, Frequency, and Monetary value. These metrics are essential for identifying valuable customers, predicting customer churn, and developing marketing plans.However, traditional RFM analysis tends to focus solely on business-related metrics and often ignores the impact of demographic characteristics such as geographic location. In this study, we extend the classic RFM approach by integrating country-specific data and acknowledge the potential impact of different regions on purchasing behavior. By integrating geographic data, we aim to explore how a customer's country can provide valuable insights to improve customer experience.This study compares two customer segmentation models, RFM-based segmentation, which uses occasion, frequency, and monetary value to group customers based on their purchasing behavior and RFM with Population-based clustering, which have both RFM variables and the customer's country as an additional demographic feature.

II. RELATED WORKS

A. RFM Clustering in Customer Segmentation

Valmohammadi (2017) [1] investigates the relationship between CRM innovation and performance using RFM analysis. By segmenting customers based on routine, frequency, and financial value, businesses can better target their businesses. However, this study also highlights the limitations of the RFM model, which does not account for

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

demographic and geographic factors that may influence purchasing decisions. This inconsistency highlights the need for advanced models that integrate geographic and market data in RFM analysis. Some other research [2] , emphasizes the importance of geographic vulnerability in understanding consumer behavior, further supporting the reason for enhanced RFM frameworks and model for clustering.

B. The Role of Demographic and Geographic Data in Customer Clustering Research by Moral (2010) [3] and Pellicone et al. (2018)

[4] researched impact of environmental changes like climate and agriculture on purchasing power of the consumers. While these research does provide valuable insights into the role of location in customer experience, they cannot be directly applied to all CRM systems. implementing financial and spatial data into CRM models can increase the accuracy of customer segmentation just by accounting for geographic differences in purchasing decisions based on the purchasing power, ultimately leading to better business outcomes. Similar studies [5] , explore the use of spatial-temporal models for analysis of customer data, reinforcing the necessity of incorporating geographic variables in the model.

III. RFM VS RFM-POPULATION MODEL

A. RFM Model (Baseline Model)

TABLE I. RFM-VARAIBLES

Variable	Meaning
Recency (R)	How recently a customer made a purchase
Frequency (F)	How often a customer buys
Monetary (M)	Total spending by a customer over time

The RFM (Recency, Frequency, Monetary) is a popular and well-defined model for segmenting customers based on their behavior. It helps businesses analyze customer spending patterns through these three basic variables.
The RFM model is very important for better marketing and customer relationship management (CRM). benefits include three major advantages.
By Recognizing high-value customers businesses can target customers with better RFM scores to maximize their revenue.
Increasing customer retention by engaging frequent customers with recent transactions increases their loyalty towards the business.
Better marketing strategies after segmenting users based on their buying habits allows the business for the betterment of marketing initiatives.

B. RFM-Population Model

TABLE II. RFMP-VARAIBLES

Variable	Meaning
Recency (R)	How recently a customer made a purchase
Frequency (F)	How often a customer buys
Monetary (M)	Total spending by a customer over time
Population (P)	Number of buyers from specific regions

Expanding on the traditional RFM model, the RFM- Population model introduces an additional aspect Population (P) to factor in geographical factors in purchasing behavior. This improvement enables businesses to analyze buying trends across different locations, leading to betterment decision-making. Integrating population data adds an additional layer to customer analysis, offering even more deeper insights. There are three key advantages for this model over the traditional model.

© 2025 IJIRCCE | Volume 13, Issue 3, March 2025| DOI: 10.15680/IJIRCCE.2025.1303088

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

IV. IMPLEMENTATION AND EVALUATION

A. Dataset

The dataset has transactional data from an online retail business based in United Kingdom specializing in unique all-occasion giftware. It includes data collected over a period of approximately more than two years, capturing data such as invoice numbers, product codes, descriptions, quantities, pricing, customer IDs, and the countries where customers are making purchase from. The business primarily serves wholesalers, and the dataset provides us insights into purchasing behavior, sales trends, and customer interactions within the given timeframe.

Dataset's comprehensive structure allows for in-depth analysis of customer purchasing patterns, purchase power based on locations, product performance, and sales trends. Additionally, since the dataset covers a business that primarily serves wholesalers, it provides insights into bulk purchasing behavior as well, which is useful for customer segmentation, inventory management, and demand forecasting.

B. Data Preprocessing

To maintain data integrity in the large datasets missing values, particularly in critical values such as Customer ID and Invoice Number, can introduce inconsistencies and inaccuracies. To remove this issue, all rows with missing Customer ID or Invoice values are removed, making sure that only complete and meaningful records remain, enhancing the reliability of analysis and preventing biases in clustering results.

Negative values in the Quantity column identifies returned product or transaction which were canceled. They can distort key customer behavior metrics, such as purchase frequency and total monetary value. Thus, all rows containing negative values are removed. This process ensures that only valid transactions are considered to the customer segmentation analysis, leading to more accurate results.

The InvoiceDate column, stored as a string, is converted into a standardized datetime format. This transformation allows for precise time-based calculations, such as recency analysis and trend detection. Proper date formatting also allows efficient aggregation and filtering based on the time intervals.

C. Feature Engineering: RFM Model

A reference date, set as one day after the most recent transaction in the dataset, is used to calculate recency with respect to date and time values. Each customer's most recent transaction date is calculated, and the difference between this date and the reference date is considered recency. This metric provides information about how recently a customer has engaged with the business.

Then by grouping the dataset by Customer ID and counting the unique invoice numbers, frequency values are calculated. These values help in identifying loyal customers who make purchases frequently over time. High-frequency customers are often more important and can be targeted for retention strategies.

Lastly monetary metric is calculated by multiplying the Quantity of each item purchased by its Unit Price and then aggregating these values for each customer. Understanding monetary contributions allows companies to differentiate high-value customers from low-spending ones.

The three key RFM attributes are then merged into a consolidated dataset. This structured representation allows clustering and segmentation analyses. The RFM model serves as the basis for identifying distinct customer segments based on their purchasing behavior. For analysis and visualization graph was plotted using matplotlib.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

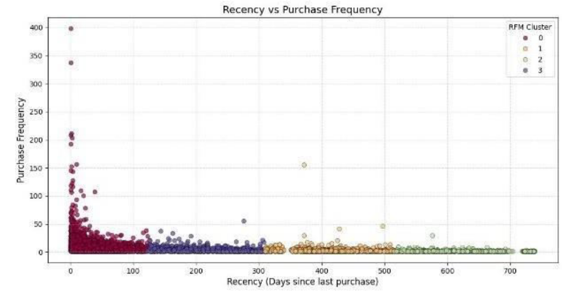


Fig. 1. RFM Frequency vs Recency

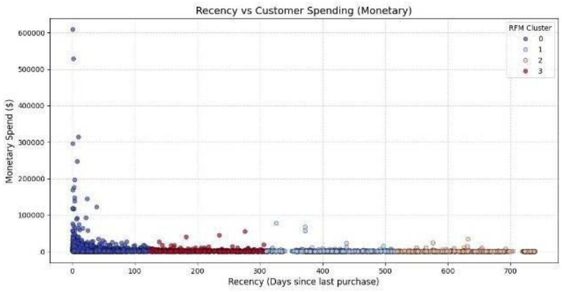


Fig. 2. RFM Monetary vs Recency

Throughout The scatter plot Figure 1 illustrates four clusters, each representing different customer purchasing behaviors.

V. RFM-POPULATION MODEL VISUALIZATION AND EVALUATION

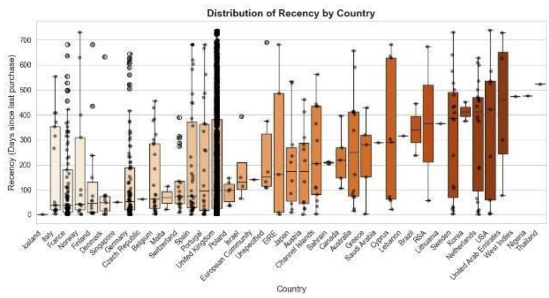


Fig. 3. Distribution of recency by country for RFMP

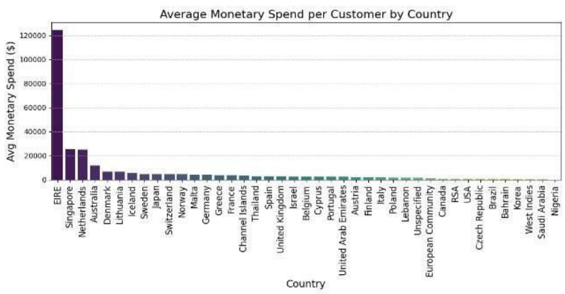


Fig. 4. Average monetart distribution according to country for RFMP

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The RFM-Population model serves as a crucial analytical framework for evaluating customer behavior by examining Recency, Frequency, Monetary while considering Population attributes across different geographical regions. This section presents an in-depth visualization and evaluation of the recency and monetary value distributions by country, providing actionable insights into consumer engagement and purchasing power of different regions.

The recency distribution analysis, as visualized by the box plot, reveals significant variations in customer activity across nations. Countries such as Brazil, Lebanon, Cyprus, Saudi Arabia, and Nigeria exhibit high median recency values, indicating a substantial proportion of inactive customers. Additionally, Brazil and Lebanon display notable outliers, suggesting the presence of long-term inactive customers who may require re-engagement strategies. In contrast, Switzerland, Denmark, Malta, and Singapore demonstrate low recency values, signifying frequent customer activity and consistent purchasing behavior. The United Kingdom (UK) and Israel present large variability, where both active and inactive customers coexist within these markets. Furthermore, regions classified under "European Communities" and "Unspecified Region" indicate a wide range of recency values, reflecting very diverse purchasing behaviors.

To complement the recency analysis, the monetary value distribution highlights key revenue-driving countries. The UK emerges as the dominant market with the highest concentration of high-spending consumers, as well as significant outliers in Singapore and Japan. These outliers denote the presence of premium consumers with exceptionally high spending tendencies. Meanwhile, countries such as Germany, Australia, and Sweden exhibit moderate spending with isolated outliers, whereas most European and Middle Eastern nations reflect lower median monetary values. The interquartile range (IQR) analysis further reveals that some nations maintain consistent spending behavior, while others, including Singapore, Japan, and the UK, exhibit wide variations, indicating diverse consumer purchasing power. From the perspective of a business, these insights facilitate strategic decision-making. High-recency countries necessitate targeted customer re-engagement campaigns, such as personalized offers and loyalty programs, to mitigate churn risks. Conversely, low-recency countries with high engagement rates warrant increased marketing investments to sustain the same growth. Additionally, the identification of high-spending outliers in Singapore and Japan suggests potential for market expansion through premium product offerings. Further analysis, such as cross-referencing recency with monetary value, can aid in segmenting high-value customers and optimized retention strategies can be formed.

The visualization and evaluation of the RFM-Population model underscore the importance of data-driven decision-making in customer relationship management. The insights derived from recency and monetary value distributions can significantly enhance targeted marketing strategies, resource allocation, and overall business profitability.

VI. CONCLUSION

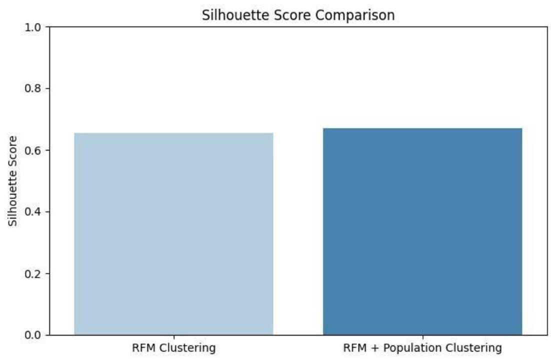


Fig. 5. Silhouette Score RFM vs RFMP

The results of our analysis demonstrate that customer segmentation using RFM analysis and clustering techniques provides valuable insights for targeted marketing strategies. The integration of KMeans clustering with RFM features alone gave a Silhouette Score of 0.653, indicating a well- formed segmentation. However, when incorporating population demographics (country), the clustering quality improved, achieving a higher Silhouette Score of 0.670. This suggests that demographic information enhances customer segmentation, leading to more accurate groupings while

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

making use for popularly used variable.

Our analysis identified key spending behaviors across many different clusters. The visualization of monetary, recency, frequency and Population distributions highlighted significant variations between different customer groups. Additionally, the country-wise analysis provided further information on understanding regional spending patterns and purchasing power.

Beside the good results, some limitations should be considered for future work. The dataset does contain biases, such as missing values and outliers, which can influence the clustering accuracy which is a big issue in RFM models. Future research could explore alternative clustering methods, such as DBSCAN and hierarchical clustering method. Moreover, incorporating some additional behavioral metrics, such as customer frequency or product category preferences, clicks on the online store, may further refine segmentation outcomes.

In conclusion, the application of RFM-based clustering, combined with demographic insights (RFMP), proves to be a robust method for customer segmentation. These findings underscore the importance of leveraging multiple data dimensions to enhance business intelligence and optimize existing marketing strategies.

VII. FUTURE WORK

While our study demonstrates the effectiveness of RFM- based clustering enhanced with demographic data and variable, several areas for future research does remains. One potential direction is advanced clustering algorithms, such as Gaussian Mixture Models (GMM) and self-organizing maps, to further improve segmentation quality for this RFM-population model. These methods could offer even better flexibility in identifying some overlapping customer segments. our study demonstrates the effectiveness of RFM- based clustering enhanced with demographic insights, several scopes for future research remains. One potential direction is the exploration of advanced clustering algorithms, such as Gaussian Mixture Models (GMM) or self-organizing maps, to further improve segmentation quality for this model. These methods could offer greater flexibility in identifying some overlapping customer segments and capturing nuanced spending behaviors.

Additionally, implementing real-time data streams could also enable dynamic segmentation, allowing businesses to adapt marketing strategies based on evolving customer behaviors with time. The integration of time-series analysis techniques may help track changes in purchasing patterns over time, providing a deeper understanding of customer lifecycle stages and with time trends.

Another promising area nowadays involves leveraging deep learning approaches for customer segmentation. Autoencoders and neural network-based clustering methods could extract latent features from high-dimensional customer data, enhancing segmentation granularity and predictive capabilities as well.

Moreover, future studies could investigate the role of contextual factors, such as seasonal trends and social media interactions, in influencing customer segmentation. By incorporating these external variables, businesses can refine their targeting strategies, improve customer engagement and make the model better.

Finally, an extension of this research could involve developing a recommendation system using collaborative filtering or reinforcement learning models for CRM. Such a system could provide personalized marketing interventions, optimizing customer retention and increasing overall profitability in CRM specific user interface.

By addressing these directions, future research can build upon our findings to further enhance customer segmentation methodologies and drive data-driven decision-making in marketing for customer relationship management.

REFERENCES

[1] C. Valmohammadi, "Customer relationshipmanagement: Innovation and performance," International Journal of Innovation Science, vol. 9, no. 4, pp. 374-395, 2017.
[2] P. Shitangsu, "Vulnerability Concepts and its Application in Various Fields: A Review on Geographical

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Perspective," Journal of Landscape and Ecological Studies, vol. 8, pp. 63-81, 2014.
[3]F. J. Moral, "Comparison of different geostatistical approaches to map climate variables: application to precipitation," International Journal of Climatology, vol. 30, 2010.
[4]G. Pellicone, T. Caloiero, G. Modica and I. Guagliardi, "Application of several spatial interpolation techniques to monthly rainfall data in the Calabria region (southern Italy)," International Journal of Climatology, vol. 38, no. 9, pp. 3651-3666, 2018.
[5]J. R. Bradley, S. H. Holan and C. K. Wikle, "Multivariate spatio- temporal models for high-dimensional areal data with application to

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details